



Regression-based, regression-free and model-free approaches for robust online scale estimation

K. Schettlinger, S. Gelper, U. Gather and C. Croux

DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

Regression-based, regression-free and model-free approaches for robust online scale estimation

Karen Schettlinger^{a,*}, Sarah Gelper^b,
Ursula Gather^a, and Christophe Croux^b

^a*Fakultät Statistik, Technische Universität Dortmund,
44221 Dortmund, Germany*

^b*Faculty of Business and Economics, Katholieke Universiteit Leuven,
3000 Leuven, Belgium*

Abstract

This paper compares methods for variability extraction from a univariate time series in real time. The online scale estimation is achieved by applying a robust scale functional to a moving time window. Scale estimators based on the residuals of a preceding regression step are compared with regression-free and model-free techniques in a simulation study and in an application to a real time series.

In the presence of level shifts or strong non-linear trends in the signal level, the model-free scale estimators perform especially well. However, the investigated regression-free and regression-based methods have higher breakdown points, they are applicable to data containing temporal correlations, and they are much more efficient.

Keywords: real-time estimation, robustness, time series, variability, volatility.

*Corresponding author. Email: schettlinger@statistik.tu-dortmund.de

1 Introduction

The variability of a time series is an important feature which helps understanding, interpreting, and forecasting complex dynamic systems in various fields of application, especially when combined with information on other significant characteristics like e.g. the location of the signal level or the direction of a trend.

In finance, for example, the variability of returns is associated with risk and thus directly relates to portfolio management and option pricing. If the volatility shows an increase, share holders should be able to react immediately to minimise their risk.

In medicine, the heart rate variability is used as a predictor for arrhythmias, for prediction of severity of illness and the mortality risk, or for choosing the right therapy for a patient. Changes in the variability of physiological variables contain useful information about the patient's state of health and thus, variability analysis is particularly useful at intensive care units where a large number of variables is measured continuously and altered conditions have to be detected online since the patients are critically ill (Seely and Macklem, 2004).

In both fields of application the time series are measured at high frequency, i.e. once per minute, once per second or in even shorter time intervals. Therefore, the methods applied for variability analysis should have a low computation time to deliver results with a minimal time delay. High frequency measurements typically lead to 'unclean' and noisy time series containing irrelevant outliers.

For high frequency financial data, outliers are often induced by the opening and closing of the markets, the release of new values for macro-economic indicators or the announcements of news. While these outliers increase the total volatility at the moment of their occurrence, we do not want them to affect the volatility estimate at further points in time, when there are no outliers present. Therefore, a robust procedure is required which estimates the smoothly varying part of the volatility, and allows to separate the 'contamination' component of the volatility from the smooth signal. This is important for volatility forecasting (Andersen, Bollerslev and Diebold, 2007).

Physiological time series from an intensive care monitoring system often contain spikes due to technical problems (e.g. like short-term disconnection of the pulse oximetry sensor) or sudden movements of the patient. Such outliers contain no information about the true state of the patient but may influence statistical analyses severely, see e.g. Charbonnier, Becq and Biot (2004). Of course, it also

may happen that outliers point at a relevant change of the patient’s health: for example, arrhythmias may appear as spikes in the heart rate series. However, such events are controlled by additional monitoring rules, e.g. making use of ECG recordings. Although these events provide crucial information on the overall state of a patient, the outliers contain no relevant information about the variability of a particular time series.

For an appropriate identification of outliers and the exact times of their appearance, additional outlier detection rules should be applied to be able to specify and explain the underlying contamination component. However, here we focus solely on the variability of a series, which should be estimated correctly in spite of the appearance of irrelevant outliers.

Such time series as discussed above usually can not be assumed to be generated by a mechanism which follows a clean model; these series are generally not stationary because they can contain trends, sudden changes in the trend and level shifts. Classical methods are not appropriate in those situations but suitable techniques are usually much less efficient.

The scale estimators considered here, are able to deal with a certain amount of outliers, with trends or even shifts in the level, they are reasonably efficient, and they are computable online i.e., they are able to present results in real time. Since we assume the global structure of the observed time series to be unknown and quite complex, we focus on moving windows techniques such that our assumptions are restricted to the local structure around a certain point in time.

In this article three different approaches for extracting the time-varying variability of a time series are compared: The first approach concerns scale estimators based on the residuals of a local linear regression; hence, we call it the *regression-based* approach. If the time series level is also of interest, the regression-based approach offers a reasonable and efficient possibility of estimation of the variability around the underlying signal level (Gather and Fried, 2003). However, if the signal level is not estimated correctly it also affects the scale estimation.

The second approach includes methods which assume the time series to be locally linear but do not require the preceding regression step for the scale estimation. We call this the *regression-free* approach. Rousseeuw and Hubert (1996) propose such estimators which are based on the vertical height of triangles formed by three observations from the sample.

Third, a *model-free* approach is considered where no assumptions on the underlying location, i.e. the time series level, are necessary and which is referred

to as the *model-free* approach. Gelper, Schettlinger, Croux and Gather (2008b) investigate such scale estimators which use the heights of triangles formed by subsequent observations. This kind of scale estimator is particularly useful when the underlying level cannot be approximated by a locally linear trend.

In this paper we compare these three different approaches by means of an extensive simulation study. Note that the regression-free estimators have not been investigated in an on-line monitoring setting yet, and there are no results on the asymptotic variances of these estimators so far.

In Section 2 the basic methodology of robust online scale estimation is described before introducing candidate methods for regression-based, regression-free and model-free scale estimation. Section 3 provides a systematic comparison of these three different types of robust online scale estimators in a simulation study where we compare (i) their finite sample efficiency, (ii) their behaviour in the presence of isolated and patchy outliers, (iii) their behaviour in the presence of level and scale shifts, (iv) their behaviour in the presence of a non-linear trend or volatility, and (v) their reaction towards temporal correlation. Section 4 shows an application to real data and finally, Section 5 provides concluding remarks and a brief outlook.

2 Methods

As a basis for the investigated methods we adopt a simple signal plus noise working model for the time series:

$$x_t = \mu_t + \varepsilon_t \text{ for } t = 1, \dots, T, \quad (1)$$

where μ_t denotes the underlying signal level and T the length of the time series, ε_t denotes an error term coming from a symmetric distribution with median zero and (time-dependent) variance σ_t^2 . This distribution may also be contaminated by outliers.

The aim of this study is to find adequate estimators of the variability of the errors ε_t without time delay, reflecting the variability of the process at time t . Suppose that the scale σ_t of the error term can be assumed to be approximately constant within a time window of length n

$$\sigma_{t-n+i} \approx \sigma_t, \quad i = 1, \dots, n. \quad (2)$$

An online scale estimation can then be achieved by applying a scale functional to the data within the most recent time window $(x_{t-n+1}, \dots, x_t)'$.

Gather and Fried (2003) compare the finite-sample performance of several high-breakdown scale functionals, including the *MAD*, the length of the shortest half (Grübel, 1988) and the S_n and Q_n scale estimators (Croux and Rousseeuw, 1992; Rousseeuw and Croux, 1993). They find these estimators to be strongly biased if the underlying signal μ_t is locally not constant. This is due to the fact that in this case the application of a scale functional to the raw data does not result in estimation of the variability of the error term alone but includes the estimation of the variability of the signal.

In the following, some scale estimators are introduced which are unbiased if the signal within one time window can be assumed to be approximately linear, i.e.

$$\mu_{t-n+i} \approx \mu_t - (n-i)\beta_t, \quad i = 1, \dots, n. \quad (3)$$

It is possible to remove the trend from the data prior to the scale estimation, e.g. by using robust regression estimators, resulting in regression-based scale estimates. Another way of dealing with such trends is the application of scale functionals to trend-adjusted data, e.g. to a sequence of first differences. However, such techniques often show only small finite sample efficiencies as e.g. the median of the absolute first differences (Gather and Fried, 2003). Here, we consider on-line scale estimators which are able to achieve reasonable efficiencies at standard normal data and which are unbiased at trended data.

For the ease of notation, we will drop the time index t in the following, such that the time window containing the n observations considered for the estimation of the most recent scale σ_t is denoted by $(x_1, \dots, x_n)'$.

2.1 Regression-based scale estimators

Gather and Fried (2003) propose to estimate the variability of the error term ε by applying a robust scale functional to the residuals $\{r_i; i = 1, \dots, n\}$ resulting from a robust linear regression fit to the current window.

In particular, they propose to use the residuals from a Repeated Median (RM) regression (Siegel, 1982) with

$$\hat{\beta}^{RM} = \text{med}_i \left\{ \text{med}_{j \neq i} \left\{ \frac{x_i - x_j}{i - j} \right\} \right\}, \quad i, j = 1, \dots, n, \quad (4)$$

$$\text{and } \hat{\mu}^{RM} = \text{med}_i \left\{ x_i + (n-i) \cdot \hat{\beta}^{RM} \right\}, \quad i = 1, \dots, n, \quad (5)$$

where the median $\text{med}\{\cdot\}$ at an even sample size n is defined as the arithmetic mean of the $(n/2)$ th and $(n/2 + 1)$ st order statistic. Here, the intercept estimate

$\hat{\mu}^{RM}$ is defined as the last fitted value within the time window, corresponding to the online estimate of the time series signal level μ . This method shows the best overall performance for online signal extraction in several comparisons (Davies, Fried and Gather, 2004; Gather, Schettlinger and Fried, 2006).

From the robust scales compared in a simulation study in Gather and Fried (2003), the Q_n scale estimation procedure (Croux and Rousseeuw, 1992; Rousseeuw and Croux, 1993) achieves quite good results. For independent Gaussian data, Q_n is more stable than other high-breakdown scale estimators: it offers an asymptotic efficiency of 82% (relative to the empirical standard deviation) which is much larger than e.g. the asymptotic efficiency of the MAD , being only 36%. Furthermore, fast computation of Q_n is possible with an update algorithm which requires $\mathcal{O}(n \log n)$ time in a worst case scenario but runs much faster in practical applications (Nunkesser, Fried, Schettlinger and Gather, 2008).

Regression-based scale estimation by Q_n based on the RM residuals $r_i^{RM} = x_i - (\hat{\mu}^{RM} - (n-i)\hat{\beta}^{RM})$, $i = 1, \dots, n$, can also be achieved in $\mathcal{O}(n \log n)$ time when using a linear time RM update algorithm (Bernholt and Fried, 2003). The corresponding Q_n scale estimator is then defined as the k th order statistic of absolute differences of RM residuals:

$$Q_n = c_n^Q \cdot \{ |r_i^{RM} - r_j^{RM}| : 1 \leq i < j \leq n \}_{(k)} \quad \text{with} \quad k = \binom{\lfloor n/2 \rfloor + 1}{2}, \quad (6)$$

where c_n^Q denotes a factor for achieving unbiasedness at normal samples of size n .

2.2 Regression-free scale estimators

Rousseeuw and Hubert (1996) propose several scale estimators based on triangular heights. Here, we adapt their approach to data observed on an equidistant design space, namely discrete time. Within the recent time window, we consider the *vertical* heights of triangles with vertices given by three observations x_i , x_j and x_k with $i, j, k \in \{1, \dots, n\}$:

$$h(i, j, k) = \left| x_j - x_i - \frac{(x_k - x_i)(j - i)}{(k - i)} \right|. \quad (7)$$

We want to point out that this vertical height does not match the 'height' defined as the perpendicular length of a triangle; the term 'height' as defined in (7) equals the non-zero residual of an L_1 regression fit to the three considered data points.

Corresponding to the idea of the Repeated Median slope estimator (4), based on slopes between pairs of observations, Rousseeuw and Hubert (1996) propose a scale estimator based on a repeated median of the heights (7) formed by triples of observations:

$$R = c_n^R \cdot \operatorname{med}_i \left\{ \operatorname{med}_{j \neq i} \left\{ \operatorname{med}_{k \neq i, j} h(i, j, k) \right\} \right\}. \quad (8)$$

As alternative to this approach they also introduce an α -quantile of the $\binom{n}{3}$ heights

$$Q_{all}^\alpha = c_n^{all} \cdot \{h(i, j, k); 1 \leq i < j < k \leq n\}_{(\lfloor \alpha \binom{n}{3} \rfloor)}. \quad (9)$$

Here, c_n^R and c_n^{all} denote again factors for achieving unbiased scale estimates at Gaussian samples of size n .

Both estimators are regression invariant but have a rather high computation time of $\mathcal{O}(n^3)$ which can be improved for the online application by using update algorithms. Furthermore, Rousseeuw and Hubert (1996) prove that the breakdown point of R is 50%.

For a scale estimator the breakdown point is determined by the minimal amount of contamination such that the estimated scale becomes either infinite (explosion) or zero (implosion). In case of an equidistant design it can be shown that for $n \rightarrow \infty$ the *explosion breakdown point* of Q_{all}^α equals $1 - \alpha^{1/3}$ while its *implosion breakdown point* corresponds to $\alpha^{1/3}$. Thus, a maximal breakdown point of 50% can be achieved when choosing $\alpha = 0.125$. However, the choice of α poses a trade-off problem between high efficiency and high robustness, or high breakdown point, respectively. Therefore, we choose a value of $\alpha = 0.5$ as a compromise in the remainder. The estimator $Q_{all}^{0.5}$ still reaches a breakdown point of approximately 21% and achieves reasonable efficiency.

The estimators (8) and (9) are regression-free in the sense that they can estimate the variability of the error term around a locally linear signal without previously estimating the trend of data via a regression fit. However, they do require that the linearity assumption (3) is valid. If this assumption is violated these methods also estimate the variability of the signal to some extent.

2.3 Model-free scale estimators

As a much faster and regression-free or even model-free version, we also consider scale estimators based on the $n - 2$ vertical heights of triangles formed by three consecutive observations x_i , x_{i+1} and x_{i+2} with $i \in \{1, \dots, n - 2\}$. For these heights, formula (7) simplifies to

$$h_i^{adj} = \left| x_{i+1} - \frac{x_i + x_{i+2}}{2} \right|. \quad (10)$$

Rousseeuw and Hubert (1996) propose a scale estimator based on an α -quantile of these adjacent heights:

$$Q_{adj}^\alpha = c_n^q \cdot \{h_1^{adj}, \dots, h_{n-2}^{adj}\}_{(\lfloor \alpha(n-2) \rfloor)}. \quad (11)$$

Gelper, Schettlinger, Croux and Gather (2008b) further propose an α -trimmed mean of adjacent heights

$$TM_{adj}^\alpha = c_n^m \cdot \frac{1}{\lfloor \alpha(n-2) \rfloor} \sum_{i=1}^{\lfloor \alpha(n-2) \rfloor} h_{(i)}^{adj} \quad (12)$$

and the square root of the α trimmed mean of squared adjacent heights

$$TMS_{adj}^\alpha = c_n^s \cdot \sqrt{\frac{1}{\lfloor \alpha(n-2) \rfloor} \sum_{i=1}^{\lfloor \alpha(n-2) \rfloor} (h_{(i)}^{adj})^2}. \quad (13)$$

Again, the values c_n^q , c_n^m , and c_n^s are factors to achieve unbiasedness at normal data and depend on the window width n .

In Gelper, Schettlinger, Croux and Gather (2008b) it is proved that the finite sample breakdown point (fsbp) for all three *adjacent-type* scale estimators (11), (12) and (13) at an empirical distribution function F_n corresponds to

$$\text{fsbp}(S_{adj}^\alpha, F_n) = \frac{1}{n} \min \left\{ \left\lceil \frac{n-1 - \lfloor \alpha(n-2) \rfloor}{3} \right\rceil, \lfloor \alpha(n-2) \rfloor \right\}.$$

Hence, the maximum breakdown point is approximately 0.25 for $\alpha = 0.25$. Furthermore, these estimators have influence functions which are smooth, symmetric around zero, and bounded for $\alpha < 1$. The efficiency becomes the larger, the larger the value of α . For the finite sample setting, it is shown in Gelper, Schettlinger, Croux and Gather (2008b) that even for small sample sizes such as $n = 20$, for various values for α the finite sample efficiencies as well as the sensitivity curve, i.e. the empirical equivalent of the influence function, are very close to their asymptotic counterparts.

As a good trade off between reasonable robustness and efficiency, Gelper, Schettlinger, Croux and Gather (2008b) recommend to apply these estimators with a value of $\alpha = 0.5$ resulting in a breakdown point of approximately 1/6 but yielding an efficiency of 40% or more at Gaussian data. The TMS_{adj}^α estimator (13) with $\alpha = 1$ is equivalent to a scale estimator proposed by Gasser and Sroka (1986) who prove that for normally distributed data $(TMS_{adj}^1)^2$ is asymptotically unbiased and strongly consistent for σ^2 . However, these properties are also true for any other value of α .

One striking advantage of the *adjacent-type* scale estimators is the low computation time: an update of any of the adjacent-type estimators can be achieved in only $\mathcal{O}(\log n)$ time. Furthermore, Q_{adj}^α , TM_{adj}^α and TMS_{adj}^α do not rely on the local linearity assumption (3) but can also estimate the variability of the error term if the signal is non-linear in any sense, e.g. if it is quadratic or contains sudden level shifts or trend changes. In this sense, the adjacent-type estimators are model-free.

3 Simulation study

In the following, we consider time series generated from the simple signal plus noise model (1) with a time series length of $T = 1000$. The simulation schemes we investigate, consist of time series with independent standard normal errors with and without contamination and of time series with autocorrelated errors. As contamination we consider isolated outliers, patches of outliers, and even level shifts. Furthermore, the effect of a sudden scale shift, the violation of the assumption of a locally constant scale and the violation of the linearity assumption for the underlying signal are examined.

For each setting $S = 1000$ time series are generated and the performance of the online scale estimators is judged at each point in time $t \in \{n, \dots, T\}$ by the mean bias

$$\text{MB}_t = \frac{1}{S} \sum_{i=1}^S \frac{\hat{\sigma}_t(i) - \sigma_t}{\sigma_t},$$

and the root mean squared error:

$$\text{RMSE}_t = \sqrt{\frac{1}{S} \sum_{i=1}^S \left(\frac{\hat{\sigma}_t(i) - \sigma_t}{\sigma_t} \right)^2}.$$

Here, n denotes the window width and $\hat{\sigma}_t(i)$ is the online scale estimation at time $t \in \{n, \dots, T\}$ for time series $i \in \{1, \dots, S\}$ where σ_t is the true (uncontaminated)

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n
$n = 20$	1.240	2.293	1.996	1.023	0.838	1.312	1.136	1.939
$n = 50$	1.221	2.427	2.094	1.023	0.824	1.301	1.145	2.092

Table 1: Finite sample consistency factors c_n for achieving unbiasedness at normal samples of size $n \in \{20, 50\}$.

scale at that time. The estimate $\widehat{\sigma}_t(i)$ represents the variability at time t in the i th simulated time series, and it is obtained by applying one of the considered scale estimators to the observations at times $(t - n + 1, \dots, t)$ in the i th time series.

For the Q_{all}^α estimator (9) and the adjacent-type scale estimators (11), (12) and (13), we consider the estimators with $\alpha = 0.5$ which offer a good compromise between high breakdown point and high efficiency (Gelper, Schettlinger, Croux and Gather, 2008b). Furthermore, the non-robust alternatives TM_{adj}^1 , the mean of all triangle heights, and TMS_{adj}^1 , the square root of the mean of squared heights, are included for comparison.

The regression-based standard deviation

$$sd = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (r_i^{LS})^2}, \quad (14)$$

calculated from least squares residuals r_i^{LS} is included as non-robust alternative, and the Q_n estimator (6) based on the RM residuals is included as robust and regression-based reference method.

In the simulations we consider the two widths $n = 20$ and $n = 50$ for the windows which are moved over the whole length of the time series, resulting in $T - n + 1$ online scale estimates for each series. We use finite sample consistency factors c_n which are derived from simulations where each of the estimators was applied to 10000 standard normal samples of size $n \in \{20, 50\}$. The factor c_n was calculated by inverting the mean of the 10000 scale estimates to achieve an unbiased estimation of the standard deviation. The applied values are given in Table 1. Furthermore, it is shown by Gelper, Schettlinger, Croux and Gather (2008a) that the finite sample consistency factor for the $Q_{adj}^{0.5}$ estimate for a sample of size n can be approximated by $c_n^q \approx 1.21 \cdot n / (n + 0.44)$.

A method performs best if it both, has a small mean bias MB_t and is efficient i.e., if it has a small $RMSE_t$, not only on average but preferably over the whole period of time.

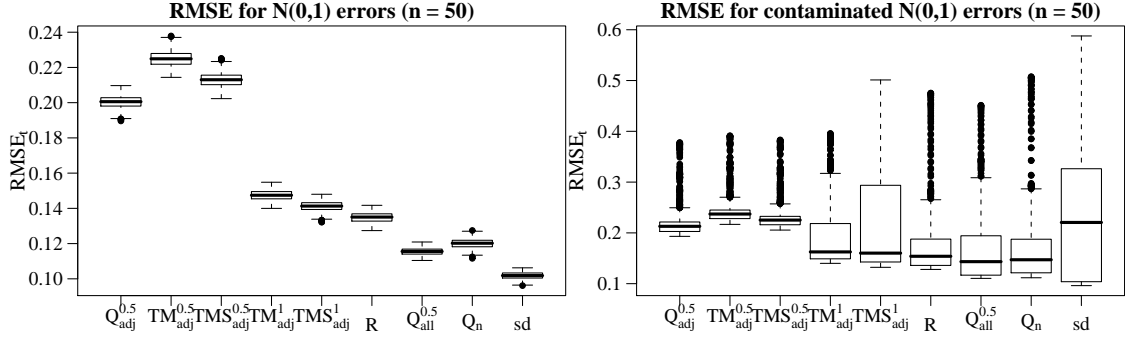


Figure 1: Boxplots of RMSE_t for time series with standard normal errors without (left) and with contamination (right) using the window width $n = 50$. The contamination comes from a $N(3, 1^2)$ or $N(0, 3^2)$ distribution, respectively.

3.1 Standard normal errors

The time series in this setting are generated according to model (1) with $\mu_t \equiv 0$ and $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, 1)$ for all $t = 1, \dots, T$.

The left panel of Figure 1 shows boxplots of the RMSE_t over all times for $n = 50$; results for $n = 20$ look similar. In this setting sd , the standard deviation based on least squares residuals (14), is, as expected, the most efficient method, showing the smallest RMSE_t . The mean of the root mean squared error, averaged over the whole observed time period,

$$\overline{\text{RMSE}} = \frac{1}{T - n + 1} \sum_{t=n}^T \text{RMSE}_t$$

can be found in Table A.1 in the appendix. The non-robust versions of the adjacent-type estimators TM_{adj}^1 and TMS_{adj}^1 are even less efficient than the robust R , $Q_{all}^{0.5}$ and Q_n . The three robust adjacent-type estimators show the largest RMSE_t with $Q_{adj}^{0.5}$ performing best from these three.

3.2 Standard normal errors with 5% contamination

To investigate the performance of the online scale estimators in the presence of contamination, we consider standard normal time series where 5% gross-error outliers are induced at fixed time points. For the first half of the time the contamination distribution corresponds to a $N(d, 1)$ -distribution leading to additive outliers of size d in the level. In the last half a $N(0, d^2)$ -distribution is used to generate scale

outliers. We consider values $d \in \{1, 2, 3, 4, 5\}$, generating small, moderately sized and large outliers. (For $d = 1$ only additive outliers are generated for $t < 500$.)

For both types of contamination five single outliers, two patches of two, two patches of three and one patch of five subsequent outliers is generated. In addition, a stretch of ten successive additive level outliers is incorporated into the simulation scheme to investigate the behaviour of the scale estimators for a level shift.

In particular, the additive outliers are induced at $t \in W_1$ and the scale outliers appear at $t \in W_2$ with

$$\begin{aligned} W_1 &= \{51, 71, 81, 86, 91, \quad 101, 102, \quad 151, 152, \\ &\quad 201, 202, 203, \quad 241, 242, 243, \quad 301, \dots, 305, \quad 401, \dots, 410\} \\ \text{and } W_2 &= \{551, 571, 581, 586, 591, \quad 601, 602, \quad 651, 652, \\ &\quad 701, 702, 703, \quad 741, 742, 743, \quad 801, \dots, 805\}. \end{aligned}$$

Figure 2 shows the mean bias averaged over time

$$\overline{\text{MB}} = \frac{1}{T - n + 1} \sum_{t=n}^T \text{MB}_t$$

as well as the root mean squared error averaged over time $\overline{\text{RMSE}}$ for the considered methods applied to a window of width $n = 50$. Plots for $n = 20$ look similar. For all methods $\overline{\text{MB}}$ and $\overline{\text{RMSE}}$ increase along with the magnitude of the outliers, determined by the value of d .

It shows clearly that the size of the outliers has a considerable influence on the bias of the non-robust scale estimators, while the mean bias does not increase as drastically with increasing d for the robust methods. Particularly for the model-free estimators $\overline{\text{MB}}$ increases the least with d , and all three model-free estimators perform very similar – better than the other methods – for all values of d .

However, in terms of efficiency these estimators perform much worse than the other robust estimators, and for small outliers i.e., small d , they even have a much higher $\overline{\text{RMSE}}$ than the non-robust methods. All methods show some increase in $\overline{\text{RMSE}}$ for increasing d , but this increase is much more pronounced for the non-robust methods. For small to moderate sizes of outliers the non-robust TM_{adj} and TMS_{adj} scales even outperform its robust counterparts in terms of $\overline{\text{RMSE}}$, but for larger sizes of outliers clearly, the robust methods are more efficient.

The right panel of Figure 1 shows boxplots of the RMSE_t evaluated at each time point t in the setting with $d = 3$ where the contamination comes either from a $N(3, 1^2)$ or a $N(0, 3^2)$ distribution. The averages $\overline{\text{RMSE}}$ for these boxplots correspond to the points at $d = 3$ in the right panels of Figure 2.

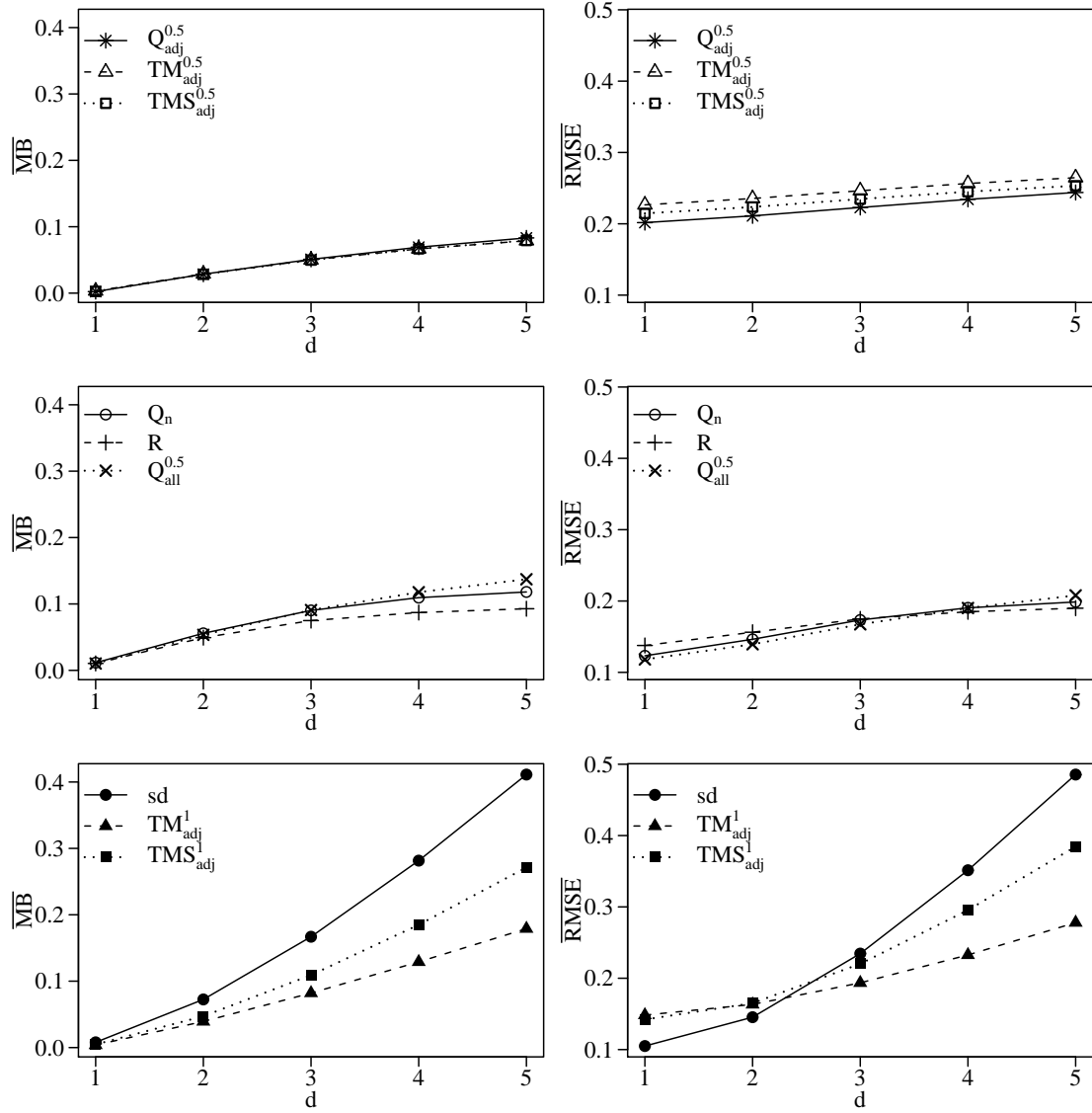


Figure 2: Average mean bias \overline{MB} (left) and average root mean squared error \overline{RMSE} (right) for the settings with contaminated normal errors where the contamination comes from a $N(d, 1^2)$ or $N(0, d^2)$ distribution, respectively, for the window width $n = 50$.

Although Figure 2 shows that the running standard deviation on average performs similar to the model-free methods, in Figure 1 it is evident that sd is heavily affected by the contamination in this setting. The TMS_{adj}^1 estimator is similarly affected, though not as bad, while TM_{adj}^1 still achieves reasonably efficient results for a non-robust estimator. Confirming this average behaviour, Figure 1 shows that from the robust methods R , Q_n and $Q_{all}^{0.5}$ perform best in terms of $RMSE_t$, with R showing the best overall performance. From the robust adjacent type estimators $Q_{adj}^{0.5}$ shows the best performance again.

To evaluate the influence of the different types and number of consecutive outliers, Figure 3 shows the mean bias MB_t over time for a normal series with contamination from a $N(3, 1^2)$ distribution for $t \leq 500$ and from a $N(0, 3^2)$ distribution for $t > 500$. For other values of d the graphics look – up to a factor – very similar. Apparently, all robust methods show some sort of reaction towards any kind of outlier but not as drastic as the non-robust scale estimators. Furthermore, there is almost no difference between the three robust adjacent-type scales.

Because all methods estimate the scale online at the right endpoint of a time window, the increase in the bias appears after the time where the outlier(s) occur. All methods, regardless whether regression-based, regression-free or model-free, show an increasing MB_t after the occurrence of outliers, a brief period of constantly increased mean bias and then a decrease; the time period until the decrease depends on the window width n .

Furthermore, it can be seen in Figure 3 that a set of single outliers results in a larger bias than consecutive ones for the adjacent-type scale estimators and the number of subsequent outliers has no impact on the magnitude of the bias. In contrast, for the scale estimators R , $Q_{all}^{0.5}$ and Q_n , requiring the local linearity of the signal, it can be observed that MB_t increases with the length of the outlier patch when additive outliers occur – most obvious for the level shift after $t = 400$.

Scale outliers seem to have a smaller influence on the MB_t for R , $Q_{all}^{0.5}$ and Q_n , while for the robust adjacent-type estimators additive outliers in the level have about the same effect as outliers in the scale. However, for the non-robust adjacent-type scale the influence of scale outliers is much worse than that of additive outliers where the non-robust estimators perform even similar to their robust counterparts.

All robust online scale estimators show a similar performance in the presence of scale outliers with R and Q_n showing the smallest MB_t for the occurrence of single outliers. On the other hand the robust adjacent-type estimators perform better for subsequent additive outliers and level shifts.

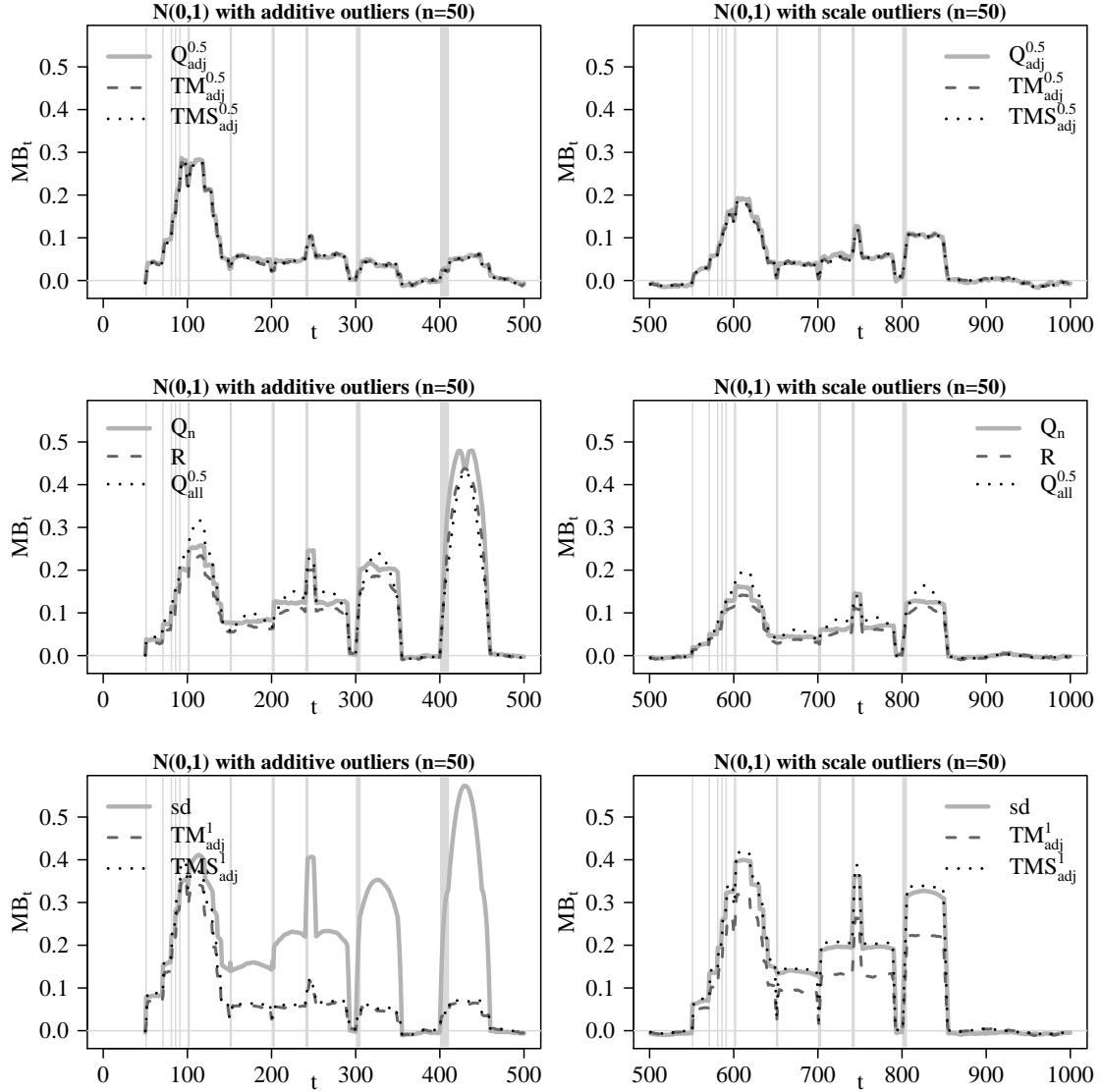


Figure 3: Mean bias MB_t over time for time series with standard normal errors and additive outliers (left) or scale outliers (right) of size five for the window width $n = 50$. The vertical grey lines mark the times where the outliers / outlier patches occur. The patch of ten outliers after $t = 400$ can also be interpreted as a level shift.

3.3 Scale shift

In this setting, the scale of the data generating normal distribution shifts from $\sigma_t = 1$ to $\sigma_t = 5$ at time $t = T/2 + 1 = 501$ while the level stays constant, $\mu_t \equiv 0$ for all $t = 1, \dots, T$.

Because of the online estimation all methods trace the scale shift with some time delay but concerning magnitude of MB_t right after the shift and the duration until the bias returns back to zero, there is no real difference between the methods. The delay basically corresponds to the chosen window width. However, in this setting it is $MB_t \approx -0.8$ right after the shift, regardless of the method or window width used (see the left panels of Figure 4).

Since all observations come from a normal distribution, the results for the root mean squared error basically correspond to the outcomes for the non-contaminated standard normal setting. Even the order of magnitude of \overline{RMSE} is quite similar to this setting, cf. Tables A.1 and A.2 in the appendix, i.e. the standard deviation sd is the most efficient method, $Q_{all}^{0.5}$, Q_n and R perform not much worse but the robust adjacent-type estimators lack some efficiency.

3.4 Slow scale change

Here, the time series are generated from a model with constant level $\mu_t \equiv 0$ for all $t = 1, \dots, T$ and independent errors from a $N(0, t^2)$ distribution i.e., the assumption of a locally constant scale (2) is violated because the scale $\sigma_t = t$ changes linearly over time.

Because all methods estimate the scale at the end of a time window, all previous observations $x_{t-n+1}, \dots, x_{t-1}$ come from distributions with a smaller variance than σ_t^2 . Therefore, all methods underestimate the true scale resulting in a negative MB_t over the whole period of time, and on average the mean bias is quite similar for all methods. Since MB_t is a measure which evaluates the bias relative to the true scale the deviation of the estimate from the true value is proportionally the largest for small values of σ_t . As the true scale σ_t increases MB_t tends to zero (see the right panels of Figure 4). When using the larger window width $n = 50$, MB_t is further away from zero than for $n = 20$.

In terms of $RMSE_t$, the results are, even in order of magnitude, similar to the standard normal and the scale shift setting (cf. Tables A.1 and A.3); that means that all investigated methods perform well even if the scale is locally not constant, for the price of a small bias.

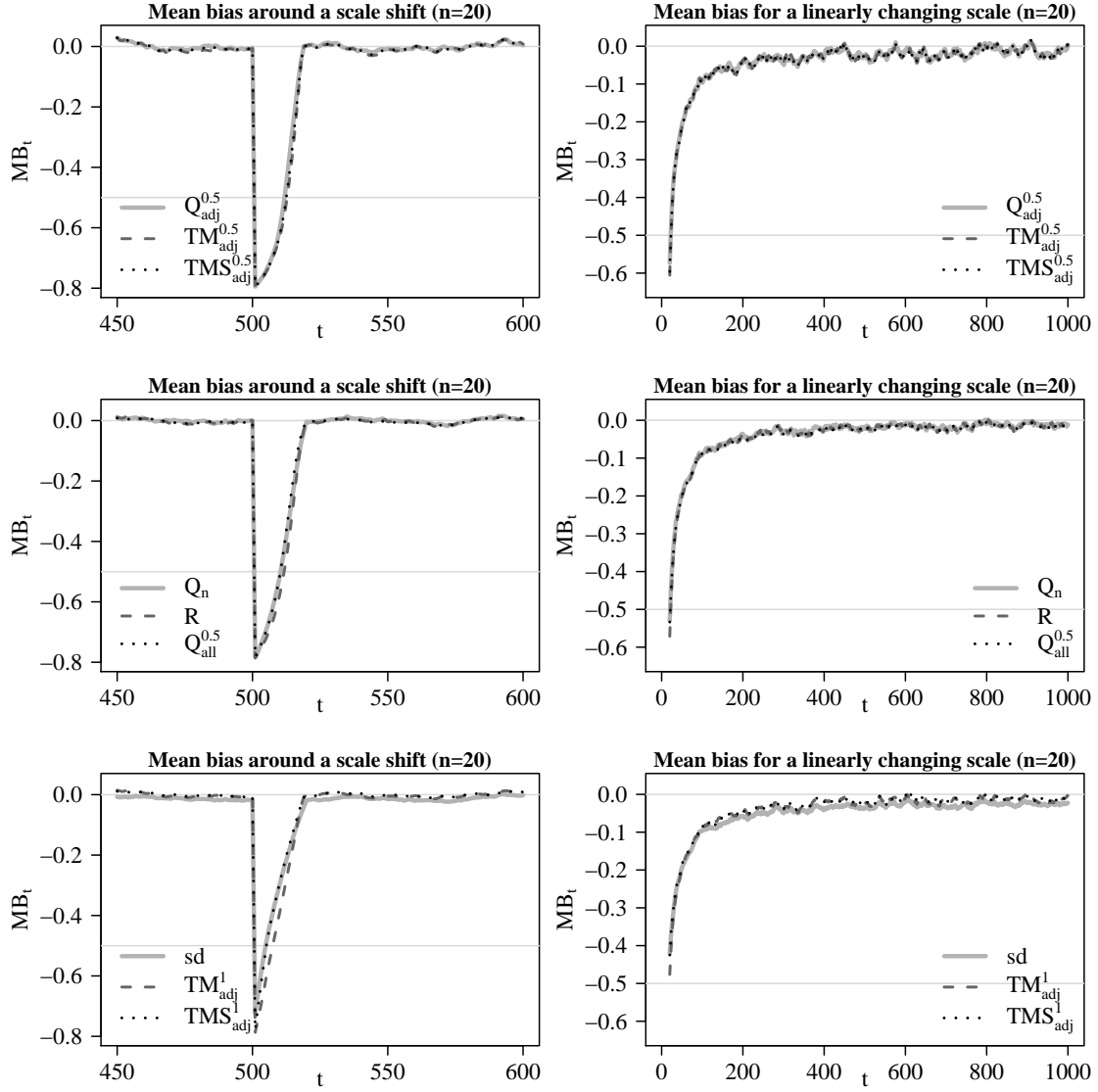


Figure 4: Mean bias MB_t for a time interval around the time $t_0 = 500$ where the scale of the independent error term shifts from $\sigma_t = 1$ to $\sigma_t = 5$ (left) and MB_t for time series with independent $N(0, t^2)$ errors (right) for the window width $n = 20$.

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.422	0.434	0.419	0.361	0.360	19.7	18.7	28.9	30.2
$n = 50$	0.385	0.400	0.390	0.333	0.318	114.3	109.1	165.1	189.0

Table 2: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with a quadratically increasing signal $\mu_t = t^2$.

3.5 Quadratic trend change

To investigate the performance of the online scale estimators when the assumption of a local linear signal within each time window is violated, we generate time series from a model with independent standard normal errors ε_t but a quadratically changing trend $\mu_t = t^2$.

In this setting MB_t and $RMSE_t$ stay almost constant over the whole period of time and $RMSE_t$ strongly depends on the underlying bias which is quite large for all methods that rely on the assumption of a local linear signal (3).

Table 2 shows the \overline{RMSE} , the average of the $RMSE_t$ values over time, for all methods and both investigated window widths. The huge difference in \overline{RMSE} between the adjacent-type scale estimators and regression-based as well as regression-free estimators emphasise the fact that the model-free scale estimators achieve very efficient estimations if there are non-linearities in the underlying signal. However, it can also be noted that the regression-free estimators R and $Q_{all}^{0.5}$ are more efficient than the regression-based Q_n and sd. TMS_{adj}^1 performs best here w.r.t. \overline{RMSE} but the robust model-free estimators do not perform much worse.

3.6 AR(1) errors

Departing from the assumption of independent errors, we consider model (1) with a constant level $\mu_t \equiv 0$ and autocorrelated errors. In particular, we generate the errors according to an AR(1) model, i.e.

$$\varepsilon_t = \varphi \varepsilon_{t-1} + e_t, \quad t \in \mathbb{Z}$$

with innovations $e_t \sim N(0, \sigma_e^2)$ for all $t = 1, \dots, T$. The unconditional variance of ε_t is then given by

$$\sigma_{\varepsilon_t}^2 = \text{Var}(\varepsilon_t) = \frac{1}{1 - \varphi^2} \sigma_e^2.$$

For the simulations we use standard normal innovations and choose the parameter for the AR(1) model $\varphi = 0.4$ for moderate correlation between successive observations which results in a marginal standard deviation of $\sigma_{\varepsilon_t} = 1.091$ for all t .

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	-0.284	-0.287	-0.286	-0.279	-0.276	-0.070	-0.095	-0.061	-0.080
$n = 50$	-0.281	-0.282	-0.282	-0.279	-0.278	-0.029	-0.043	-0.026	-0.033

Table 3: Average mean bias \overline{MB} of the online scale estimators for time series with AR(1) errors.

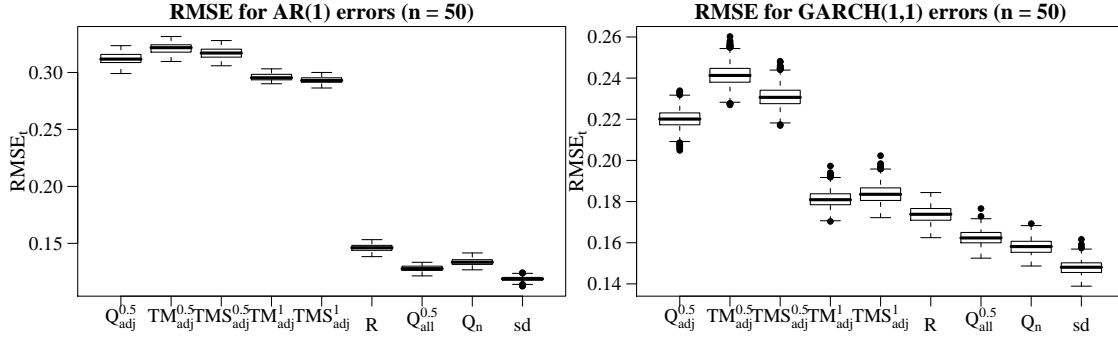


Figure 5: Boxplots of $RMSE_t$ for time series with AR(1) errors (left) and GARCH(1,1) errors (right) using the window width $n = 50$.

In case of AR(1) errors all methods loose some efficiency compared to the standard normal setting, in particular when using the larger window width (cf. Table A.4). However, the results for the RMSE are still similar to the situation of independent $N(0, 1)$ errors, only TM_{adj}^1 and TMS_{adj}^1 are much less efficient here. This can also be seen when comparing the boxplots for the $RMSE_t$ values in case of AR(1) errors in the left panel of Figure 5 with those in the standard normal situation, shown in the left panel of Figure 1.

The increase in $RMSE_t$ is partly due to the fact that because of the positive autocorrelation all scale estimators show a negative bias (see Table 3). The regression-based and regression-free methods are only a little biased, and Q_n is the least biased method. The model-free estimators have a larger bias which approximately has the same magnitude for all estimators of this type, independent of the window width $n = 20$ or $n = 50$ or of the fact whether they are robust or not. Thus, we do not recommend to use model-free estimators on time series where autocorrelations can be expected.

3.7 GARCH(1,1) errors

As a further deviation from our model assumptions we investigate time series with autocorrelated errors and a slowly varying scale, i.e. in model (1) we consider $\mu_t \equiv 0$ and ε_t following a GARCH(1,1) model with parameters $\alpha_0, \alpha_1, \beta_1 \in \mathbb{R}$

$$\begin{aligned}\varepsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.\end{aligned}$$

The mean bias MB_t and root mean squared error $RMSE_t$ are calculated w.r.t. the conditional σ_t in the uncontaminated model.

The right panel of Figure 5 shows boxplots of the $RMSE_t$ values for each of the considered methods when applying the methods to a GARCH(1,1) process using the window width $n = 50$. Similar results are achieved for $n = 20$, only that Q_n loses some efficiency while TMS_{adj}^1 gains some when using this smaller window width.

Figure 5 shows that the running standard deviation is the most efficient method for autocorrelated errors, but the robust methods Q_n , $Q_{all}^{0.5}$ and R are also quite good and even more efficient than TM_{adj}^1 and TMS_{adj}^1 . The three robust adjacent-type estimators are least efficient for GARCH(1,1) errors whereas $Q_{adj}^{0.5}$ is the best one from these three. Table A.5 contains the corresponding \overline{RMSE} values, showing that compared to the $N(0, 1)$ setting (cf. Table A.1) there is a loss in efficiency for all methods although the results are quite similar in both situations.

For GARCH(1,1) errors the bias of all methods is much smaller than in the AR(1) setting, and it is always close to zero. However, just like in the AR(1) setting, a considerable negative bias can be observed for the robust model-free scale estimators. This appears even stronger for a larger value of n , while a larger window width improves the bias for sd , Q_n , $Q_{all}^{0.5}$, and R , i.e. all estimates which require that the signal can be locally approximated by a line.

All investigated methods provide sensible information on the slowly varying conditional scale although they estimate the local unconditional scale. However, the model-free estimators do not cope as well with the autocorrelation as the other methods.

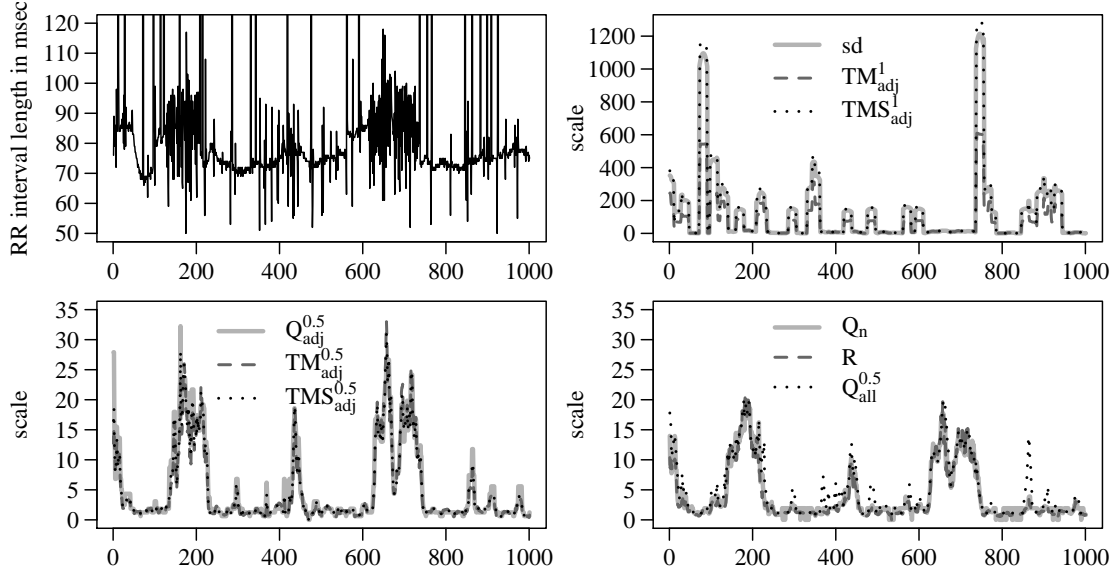


Figure 6: Sequence of 1000 RR interval lengths in msec (top left panel) with corresponding online scale estimates based on the window width $n = 20$. The top right panel shows the non-robust scale estimates, the bottom panels display the robust scales.

4 Application

In intensive care, the heart rate in beats per minute is derived from the continuous electrocardiogram (ECG) signal. However, to prevent an intermediate calculation step, the heart rate variability is generally evaluated not from the heart rate measurements which are updated once a second but from the ECG itself: Therefore, the length of the time intervals between consecutive heart beats, conventionally named RR intervals, is measured in milliseconds and the heart rate variability is calculated from a sequence of such lengths. Technically, such a sequence is not a time series, because the sampling period is not equidistant. It would be easy to convert this sequence to a time series measured e.g. once per second, where an observation corresponds to the length of the last RR interval. However, the difference to the sequence of true RR interval lengths would be minimal while there would be some loss in information and also in computation time.

The top left panel of Figure 6 shows a sequence of 1000 RR interval lengths in milliseconds from a patient taking part in the Cardiac Arrhythmia Suppression Trial (CAST). The data set is obtained from the Interbeat (RR) Interval Database on PhysioNet (Goldberger et al., 2000). It can be seen that the interval lengths

do not permanently vary around a constant level but also show trends and trend changes (see e.g. the observations from 40 to 80 and the change thereafter). For presentational reasons positive outliers with a size larger than 120msec are cut off. In this stretch of 1000 observations there are 25 such measurements with their values ranging between 653 and 5363, meaning a total percentage of 2.5% of very large outliers. However, some smaller values might also be considered as outliers because of the generally smaller variability of the observations at the time of observation, see e.g. the observations x_{811} and x_{812} .

Some very large observations are given by $x_{72} = 4838$ and $x_{97} = 2032$, as well as by $x_{737} = 5363$ and $x_{755} = 1185$. For the windows including these observations, it is quite obvious that the non-robust scale estimators, shown in the top right panel of Figure 6, 'break down' because they reach huge values of more than 1000. It can also be seen that, according to the simulation results shown in the previous section, from the non-robust scale estimators the mean of adjacent triangle heights TM_{adj}^1 does not show such a strong reaction to outliers as TMS_{adj}^1 and the standard deviation. This appears even more obvious when using a larger window width like $n = 50$ (not shown here).

In contrast, the robust online scale estimators displayed in the bottom panels of Figure 6 perform well in the presence of outliers and – from looking at the data – also yield sensible estimations of the true scale. All robust methods indicate the increased variability from about 130 to 210 and from about 610 to 730 and also after 400.

The outcomes from the three robust scale estimators based on adjacent triangle heights are quite similar again, only that in times of increased variability $Q_{adj}^{0.5}$ tends to larger estimations than $TM_{adj}^{0.5}$ and $TMS_{adj}^{0.5}$. Furthermore, they all show some sort of reaction if several outliers occur within the window used for the estimation, especially if the window width is as small as, say, $n = 20$. This is due to their breakdown point of about $1/6$.

From the regression-free estimators relying on the local linearity assumption (3) the $Q_{all}^{0.5}$ scale estimator performs similar to the adjacent-type estimators while R is more robust towards the outliers and also more efficient because the R estimations do not show such a large variability. The robust Q_n scale estimator based on the RM residuals performs similarly good as R but shows larger variability.

5 Conclusions

The robust methods compared in this article provide useful tools for the online extraction of time varying variability. Because of the moving window approach all estimators have a computation time which allows for real time application and they have given proof of their usefulness at a real data example. The methods are able to deal with contaminated data, trends and trend changes and work well even if the true scale cannot be assumed as locally constant within one time window, or if the errors are autocorrelated.

From the model-free adjacent-type scale estimators the two non-robust estimators TM_{adj}^1 and TMS_{adj}^1 are reasonably efficient at standard normal data and perform much better than the standard deviation in the presence of contamination. However, outliers cause considerable bias.

The three robust model-free estimators perform very similar in all settings but $Q_{adj}^{0.5}$ is slightly more efficient than $TM_{adj}^{0.5}$ and $TMS_{adj}^{0.5}$. The advantages of the model-free scale estimators consist in their very low update computation time of only $\mathcal{O}(\log n)$ and the fact that they work well in the presence of non-linearities in the signal level such as quadratic trends or level shifts. In case of a linear change in the scale they also perform reasonably well, but for autocorrelated errors they show some bias and loss of efficiency.

Compared to the model-free estimators the investigated regression-free methods, R and $Q_{all}^{0.5}$, have higher breakdown points and are much more efficient. Furthermore, their performance in the presence of autocorrelated errors is similar to the case of independent errors. If no contamination is present $Q_{all}^{0.5}$ is slightly more efficient than R , but in comparison R shows a better overall performance because of its higher robustness against outliers.

The Q_n estimator based on the residuals from a local repeated median regression is also highly robust and efficient (even at non-contaminated data) and it has a much lower computation time than the regression-free estimators. It also provides good estimations for autocorrelated errors. However, this scale estimator strongly relies on the underlying signal estimation which can cause a large bias for the scale if estimated wrongly.

If only a low percentage of additive or level outliers can be expected and the signal level of the time series is likely to contain many trend changes and level shifts and / or if the computing time is very limited, the model-free scale estimator $Q_{adj}^{0.5}$ should be chosen for online scale estimation with a medium-sized window

width, yielding a good compromise between robustness and efficiency. However, if temporal correlations are expected or if outliers are very likely to occur, a more robust and more efficient method like R or Q_n should be applied with a window width as large as possible but still ensuring the local linearity of the signal level within this window.

Of course, the comparisons presented here are not exhaustive and further comparisons with other robust scale estimators should be performed. Furthermore, the biggest disadvantage of the highly robust and efficient R scale estimator could be overcome by a computationally efficient update algorithm.

An important issue which has not been raised in this paper is the choice of the window width which of course is of crucial importance for the scale estimation. While the model-free scale estimators seem to work better with small window widths, larger window widths improve efficiency and robustness for the regression-free and regression-based scales. However, since we do not know the time series structure beforehand, it is not possible to determine an optimal window width for a real time application and thus, the window width has to be determined from an application-oriented background or adaptively. Therefore, the development of a data-driven choice of the window width is also an important topic for future research.

Acknowledgements

We gratefully acknowledge the financial support of the German Science Foundation (DFG, SFB 475 "Reduction of Complexity for Multivariate Data Structures"), the European Science Foundation (ESF-network SACD "Statistical Analysis of Complex Data with Robust and Related Statistical Methods") and the Flemish Science Foundation FWO through project G.0594.05.

Appendix: Complementary Tables

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.325	0.360	0.341	0.240	0.229	0.230	0.194	0.210	0.167
$n = 50$	0.200	0.225	0.213	0.147	0.141	0.135	0.115	0.120	0.102

Table A.1: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with standard normal errors.

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.331	0.365	0.347	0.244	0.233	0.237	0.200	0.217	0.171
$n = 50$	0.219	0.244	0.232	0.162	0.153	0.155	0.135	0.139	0.115

Table A.2: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with normal errors and a scale shift from $\sigma_t = 1$ to $\sigma_t = 5$.

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.321	0.354	0.336	0.240	0.230	0.232	0.198	0.213	0.174
$n = 50$	0.215	0.236	0.226	0.168	0.161	0.161	0.145	0.147	0.132

Table A.3: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with normal errors and a linearly changing scale $\sigma_t = t$.

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.360	0.377	0.368	0.322	0.316	0.238	0.206	0.225	0.192
$n = 50$	0.312	0.321	0.317	0.296	0.293	0.146	0.128	0.134	0.119

Table A.4: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with AR(1) errors.

	$Q_{adj}^{0.5}$	$TM_{adj}^{0.5}$	$TMS_{adj}^{0.5}$	TM_{adj}^1	TMS_{adj}^1	R	$Q_{all}^{0.5}$	Q_n	sd
$n = 20$	0.326	0.358	0.341	0.249	0.244	0.243	0.210	0.222	0.183
$n = 50$	0.220	0.241	0.231	0.181	0.184	0.174	0.162	0.158	0.148

Table A.5: Average root mean squared error \overline{RMSE} of the online scale estimators for time series with GARCH(1,1) errors.

References

- Andersen, T.G., Bollerslev, T., Diebold, F.X. (2007) Roughing It Up: Including Jump Components in the Measurement, Modelling and Forecasting of Return Volatility. *The Review of Economics and Statistics* **89**, 701–720.
- Bernholt, T., Fried, R. (2003) Computing the Update of the Repeated Median Regression Line in Linear Time. *Information Processing Letters* **88** (3), 111–117.
- Charbonnier, S., Becq, G., Biot, L. (2004) On-Line Segmentation Algorithm for Continuously Monitored Data in Intensive Care Units. *IEEE Transaction on Biomedical Engineering* **51** (3), 484–492.
- Croux, C., Rousseeuw, P.J. (1992) Time-Efficient Algorithms for Two Highly Robust Estimators of Scale. *Computational Statistics* **1**, 411–428.
- Davies, P.L., Fried, R., Gather, U. (2004) Robust Signal Extraction for On-Line Monitoring Data. *Journal of Statistical Planning and Inference* **122**, 65–78.
- Gasser, T., Sroka, L. Jennen-Steinmetz, C. (1986) Residual Variance and Residual Pattern in Nonlinear Regression. *Biometrika* **73** (3), 625–633.
- Gather, U., Fried, R. (2003) Robust Estimation of Scale for Local Linear Temporal Trends. *Tatra Mountains Mathematical Publications* **26**, 87–101.
- Gather, U., Schettlinger, K., Fried, R. (2006) Online Signal Extraction by Robust Linear Regression. *Computational Statistics* **21** (1), 33–51.
- Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2008a) Addendum to Robust Online Scale Estimation in Time Series: A Model-Free Approach. *Technical Report, KU Leuven* <http://www.econ.kuleuven.be/sarah.gelper/public>.
- Gelper, S., Schettlinger, K., Croux, C., Gather, U. (2008b) Robust Online Scale Estimation in Time Series: A Model-Free Approach. *Journal of Statistical Planning and Inference* To appear.
- Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E. (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Sig-

- nals. *Circulation* **101** (23), e215–e220, circulation Electronic Pages: <http://circ.ahajournals.org/cgi/content/full/101/23/e215>.
- Grübel, R. (1988) The Length of the Shorth. *Annals of Statistics* **16** (2), 619–628.
- Nunkesser, R., Fried, R., Schettlinger, K., Gather, U. (2008) Online Analysis of Time Series by the Q_n Estimator. *Computational Statistics and Data Analysis* To appear.
- Rousseeuw, P.J., Croux, C. (1993) Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association* **88** (424), 1273–1283.
- Rousseeuw, P.J., Hubert, M. (1996) Regression-Free and Robust Estimation of Scale for Bivariate Data. *Computational Statistics and Data Analysis* **21** (1), 67–85.
- Seely, A.J.E., Macklem, P.T. (2004) Complex Systems and the Technology of Variability Analysis. *Critical Care* **8**, 367–384.
- Siegel, A.F. (1982) Robust Regression Using Repeated Medians. *Biometrika* **69** (1), 242–244.